

Automated Latent Fingerprint Identification Technologies

Mike Indovina

23 September 2009

Outline

- Introduction
- Problems and Issues
- Technologies & Solutions



Stakeholders

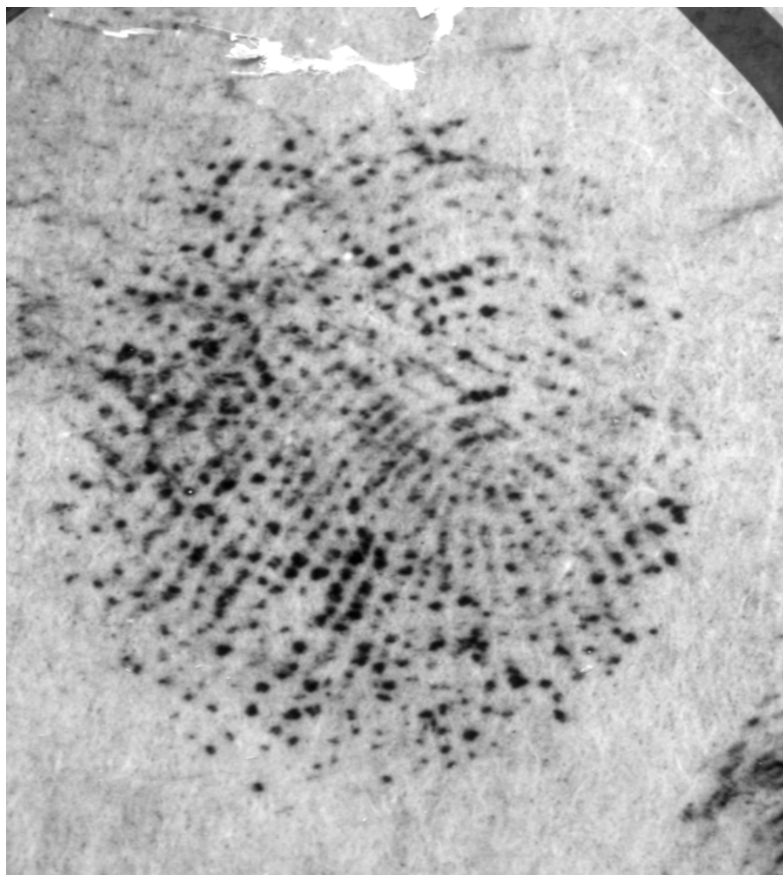
- Law enforcement (FBI, local LE, Interpol)
 - Identify perpetrators (often recidivists) of crimes
- Counter-terrorism (DHS & NCTC)
 - Real-time screening (“needles in a haystacks”)
 - “Connect the dots” (tracking unknowns)
- Forensics community at large
 - Standard definitions and practices, Daubert
 - 2009 National Academies Report



Latent Fingerprints are *Different*

- **Collected** using forensic techniques – **not** captured
- Typically lower quality than conventional fingerprints
 - More noise, distortion, “background” interference
 - Less friction ridge area
 - Non-overlapping regions (tips & sides, lower joints)
- ...So *Human assisted matching is usual practice*

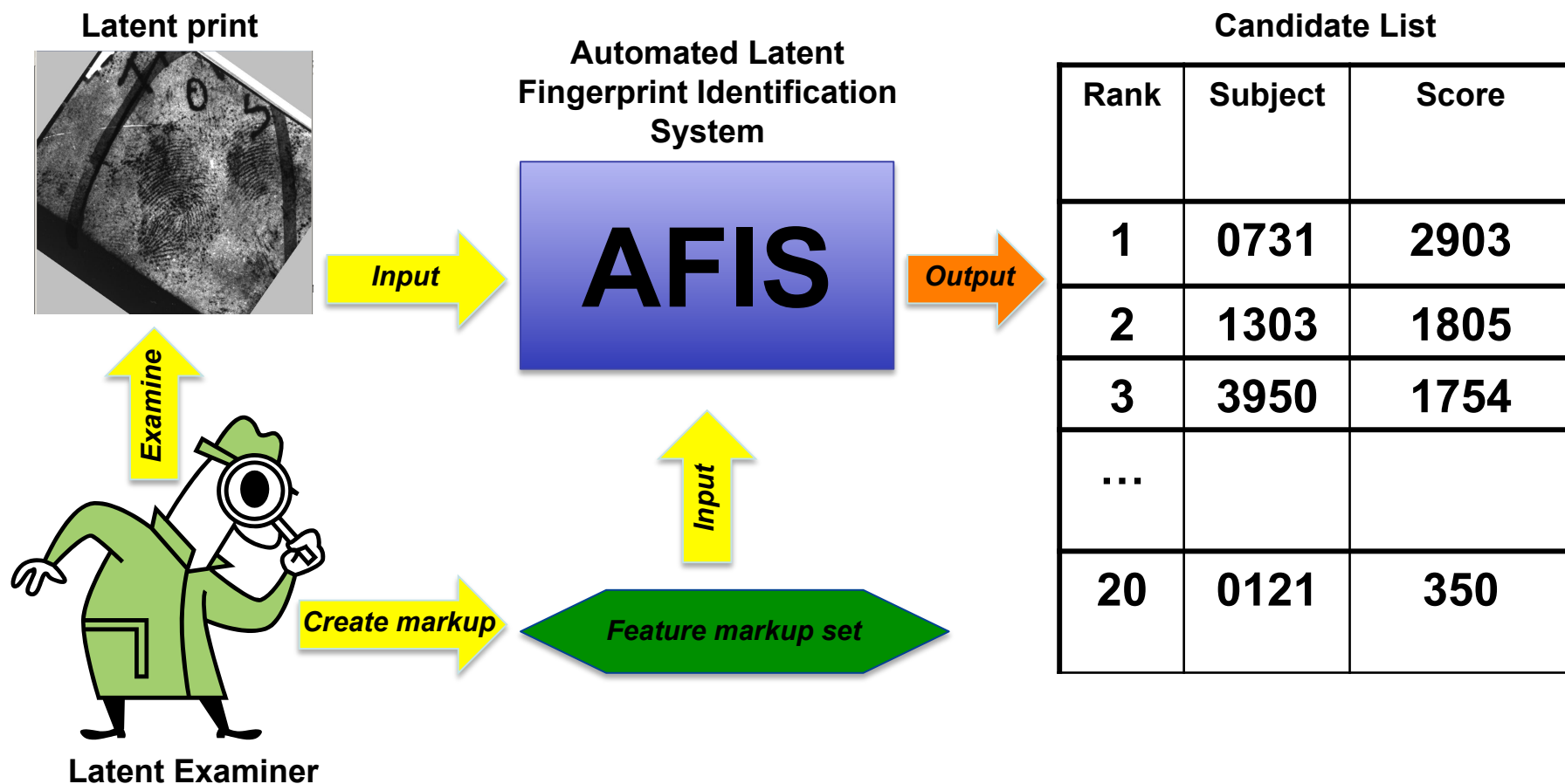
Latent Fingerprint



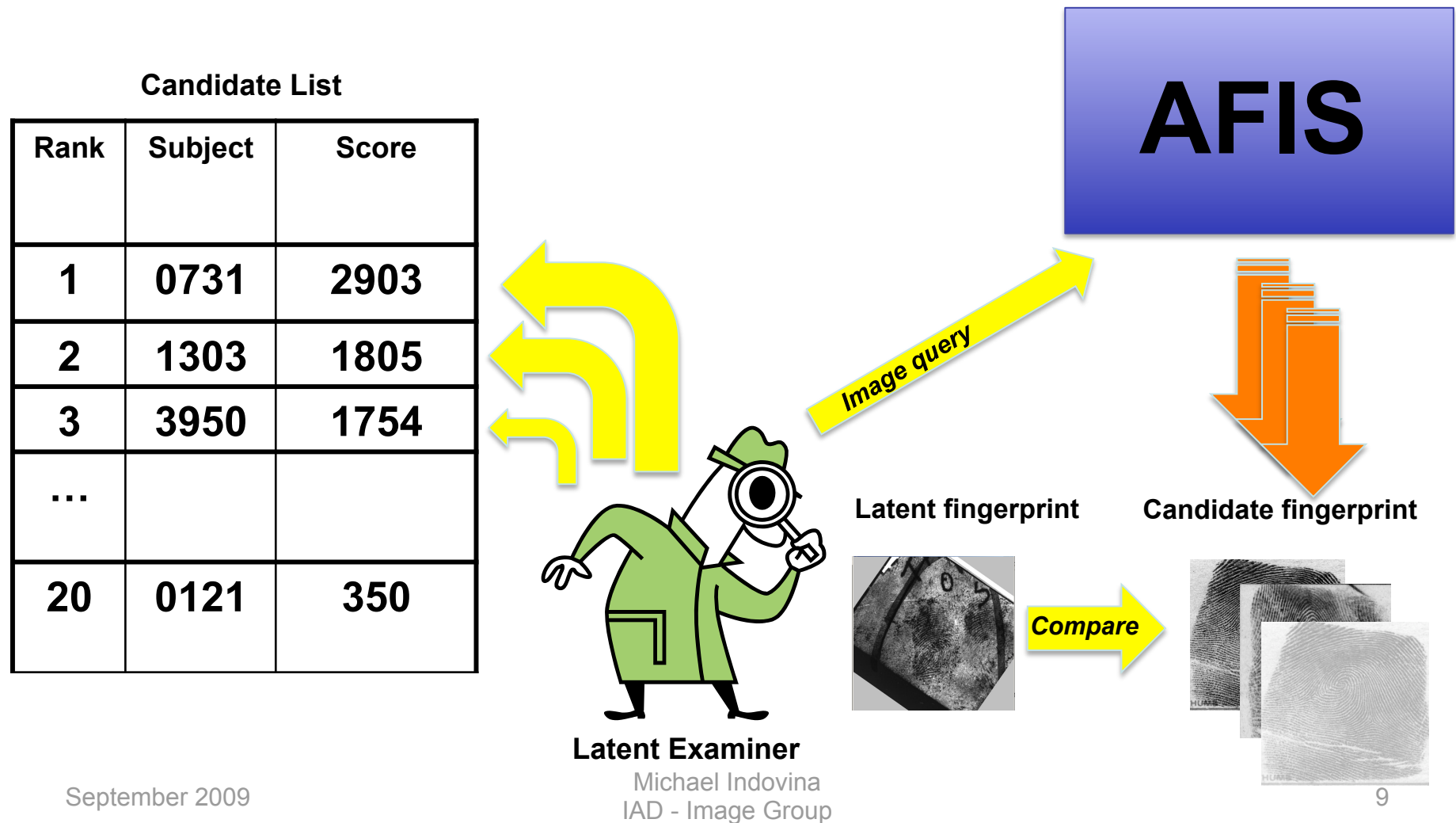
Enrolled Fingerprint



How Latent Identification works – 1



How Latent Identification works – 2



“AFIS” = manual processing required?

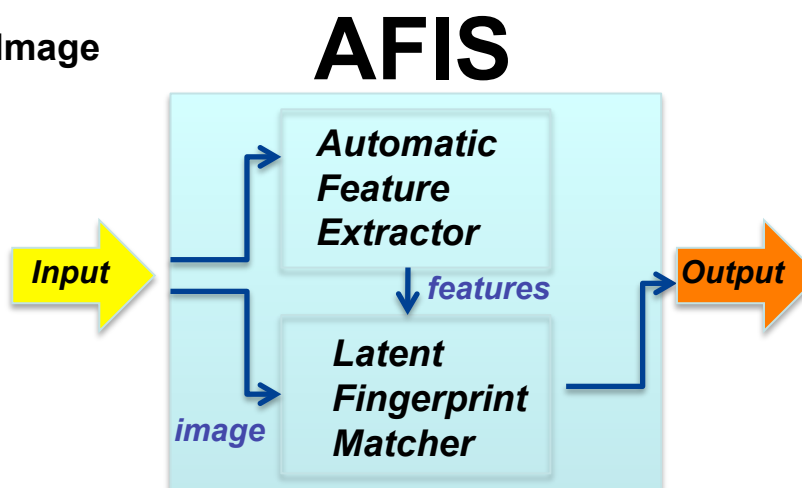
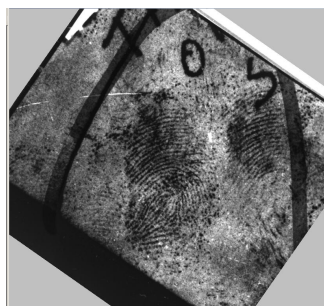
- “Front end” manual processing
 - *Image adjustment (cropping, orientation, ...)*
 - Feature selection and markup
- “Back end” manual processing
 - Individual candidate retrieval from AFIS
 - 1-to-1 comparisons: latent vs. candidate
 - Final match decisions: yes? no? inconclusive?

Problems

- Manual processing by trained personnel
 - Bottleneck to scalability and real-time operation
 - They're only human! (error, biases, variability....)
- Feature sets
 - Feature sets vary from system to system
 - Standards don't cover full range of examiner features
- Candidate lists
 - Ranking and scoring vary from system to system

Automatic Feature Extraction and Matching (AFEM)

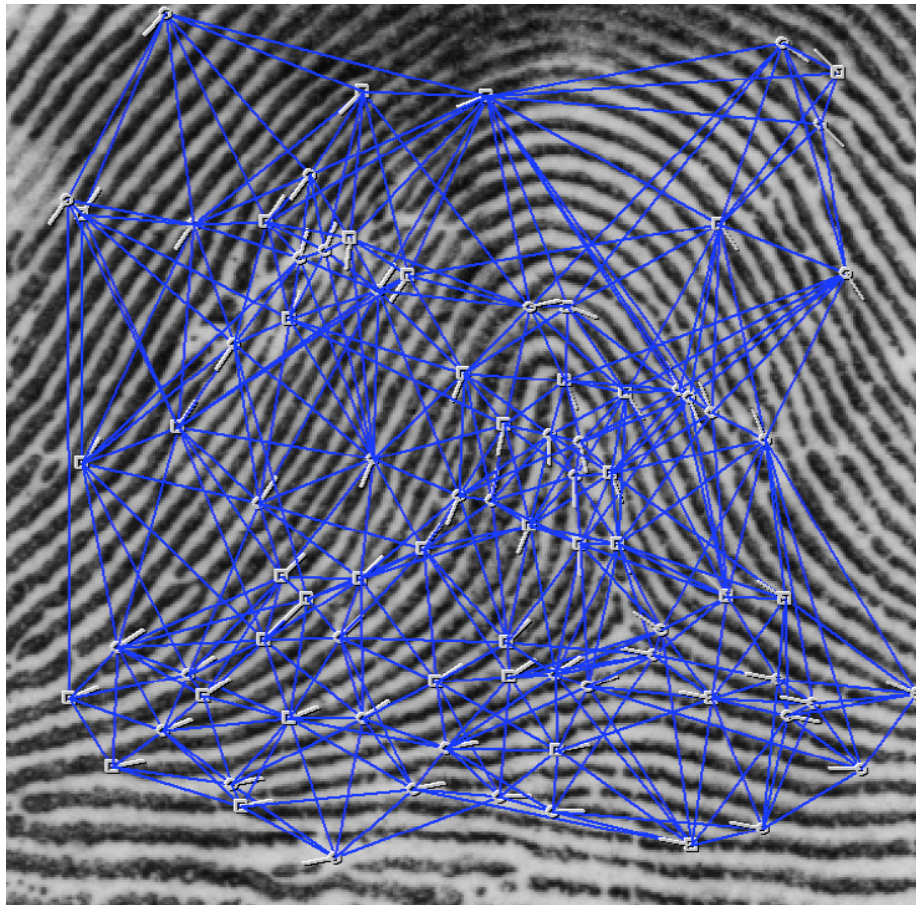
Latent Fingerprint Image



Candidate List

Rank	Subject	Score
1	0731	2903
2	1303	1805
3	3950	1754
...		
20	0121	350

Why study AFEM?



- Less manual feature selection (manual selection only when absolutely necessary)
- Images inherently more interoperable than features (*no standardized features required*)
- Algorithms are more consistent with other algorithms
- AFEM does exist but independent performance evaluation is needed to determine when to use it

Information Technology Laboratory

Information Access Division (IAD)

NIST

National Institute of
Standards and Technology



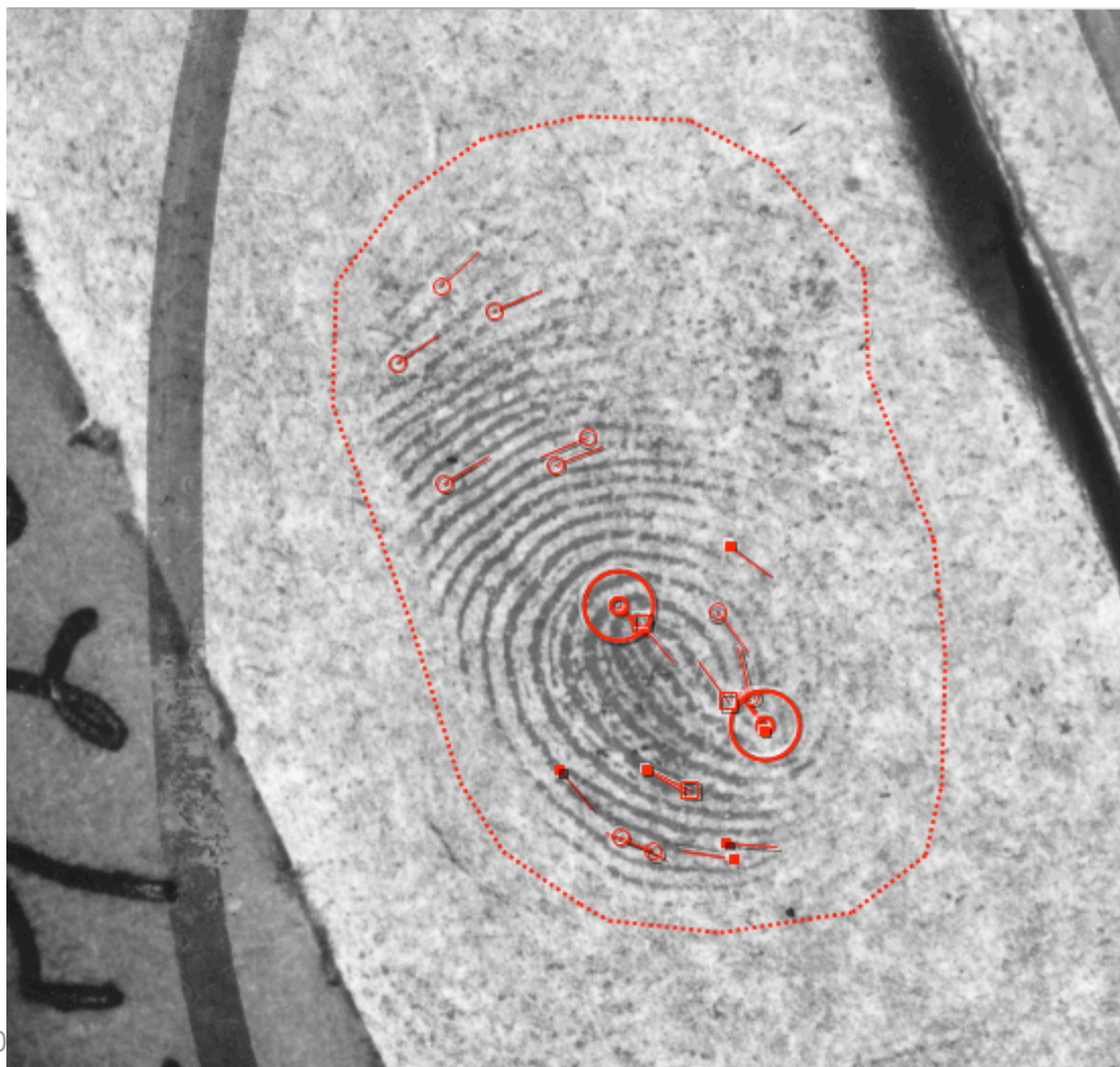
Fingerprint Feature Sets

➤ Traditional

- Traditional Feature sets ~ interoperable features
- FBI EBTS specification built on ANSI/NIST standard

➤ Proprietary

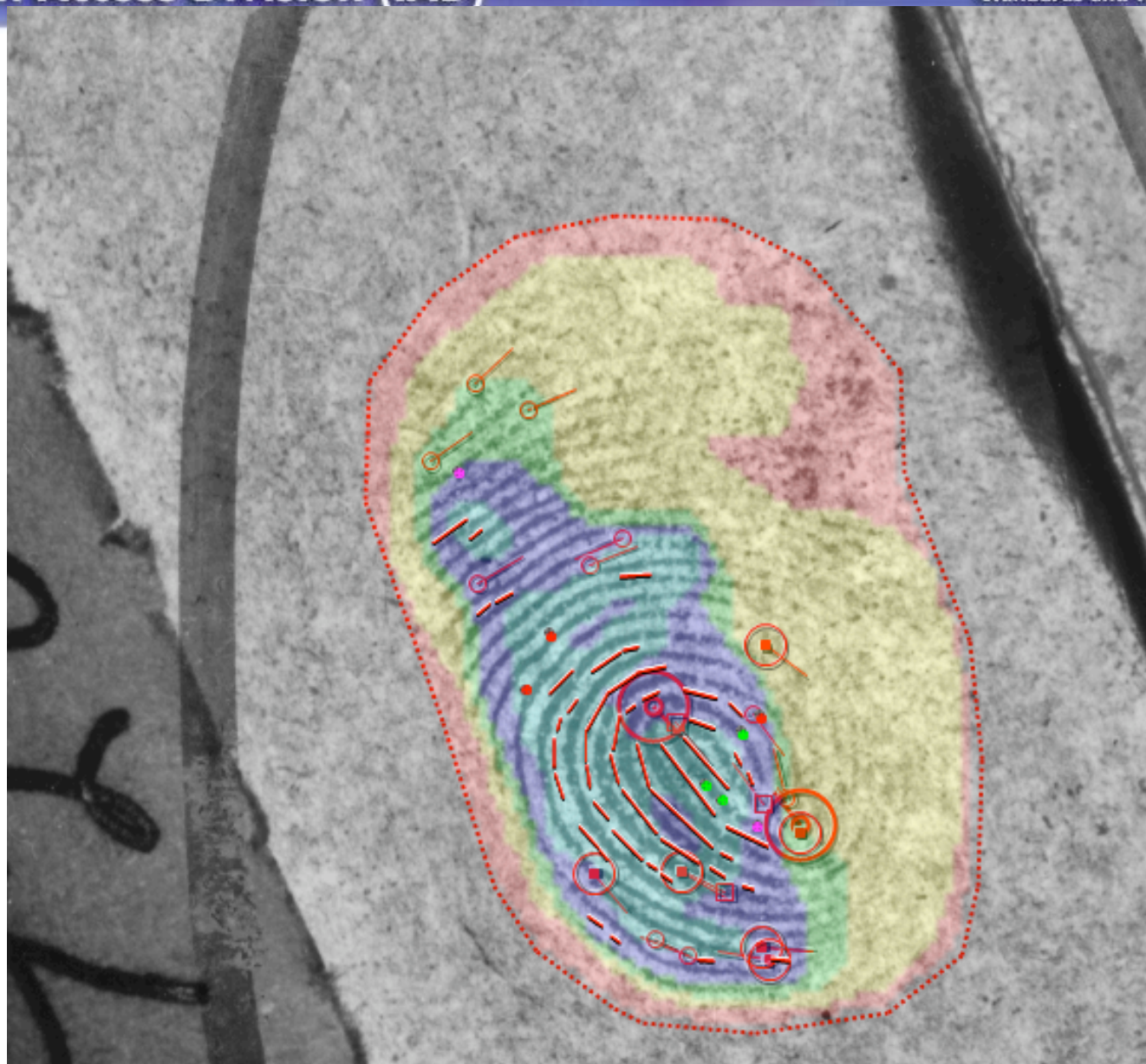
- Traditional and Proprietary feature sets
- More broadly defined
- Interoperable achieved with traditional features

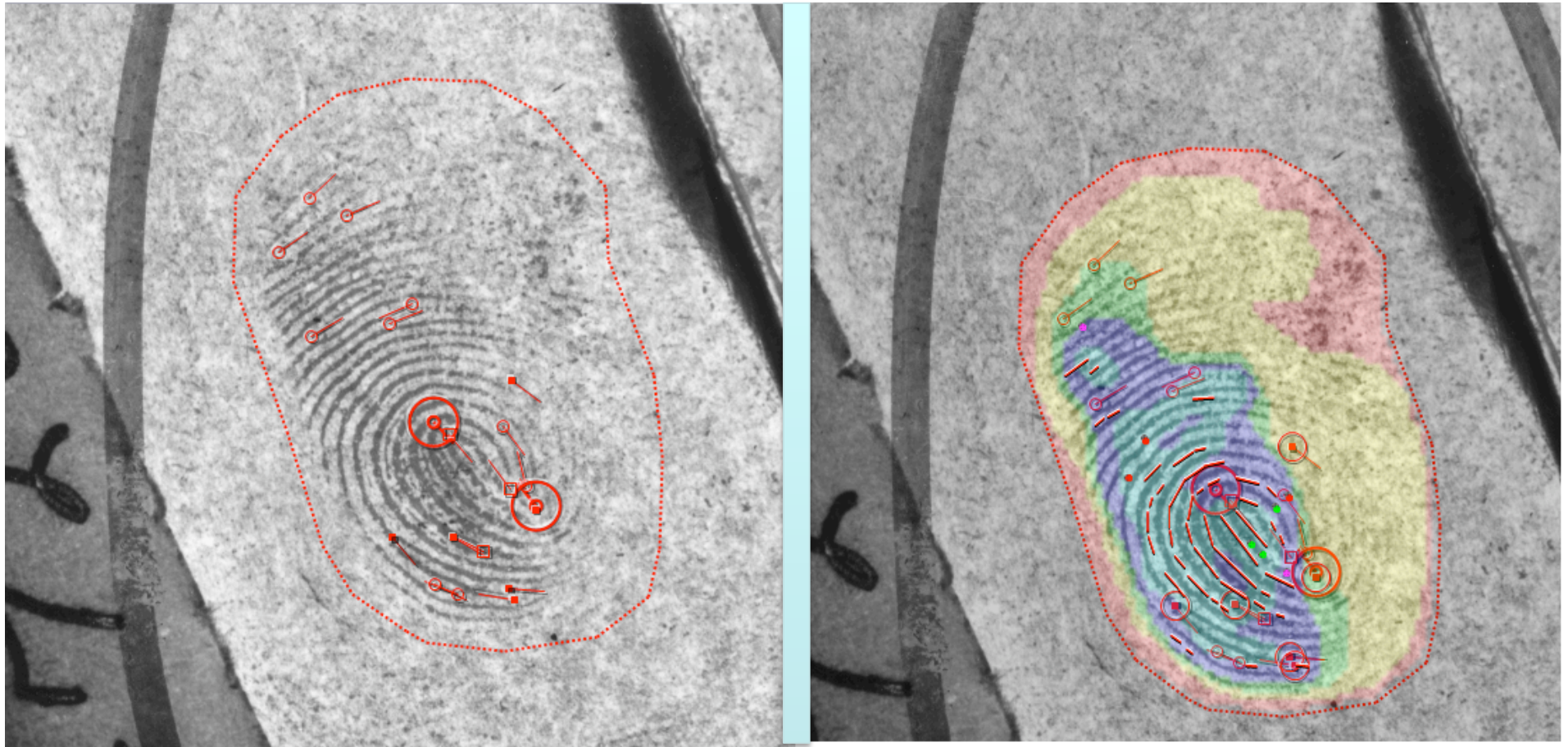


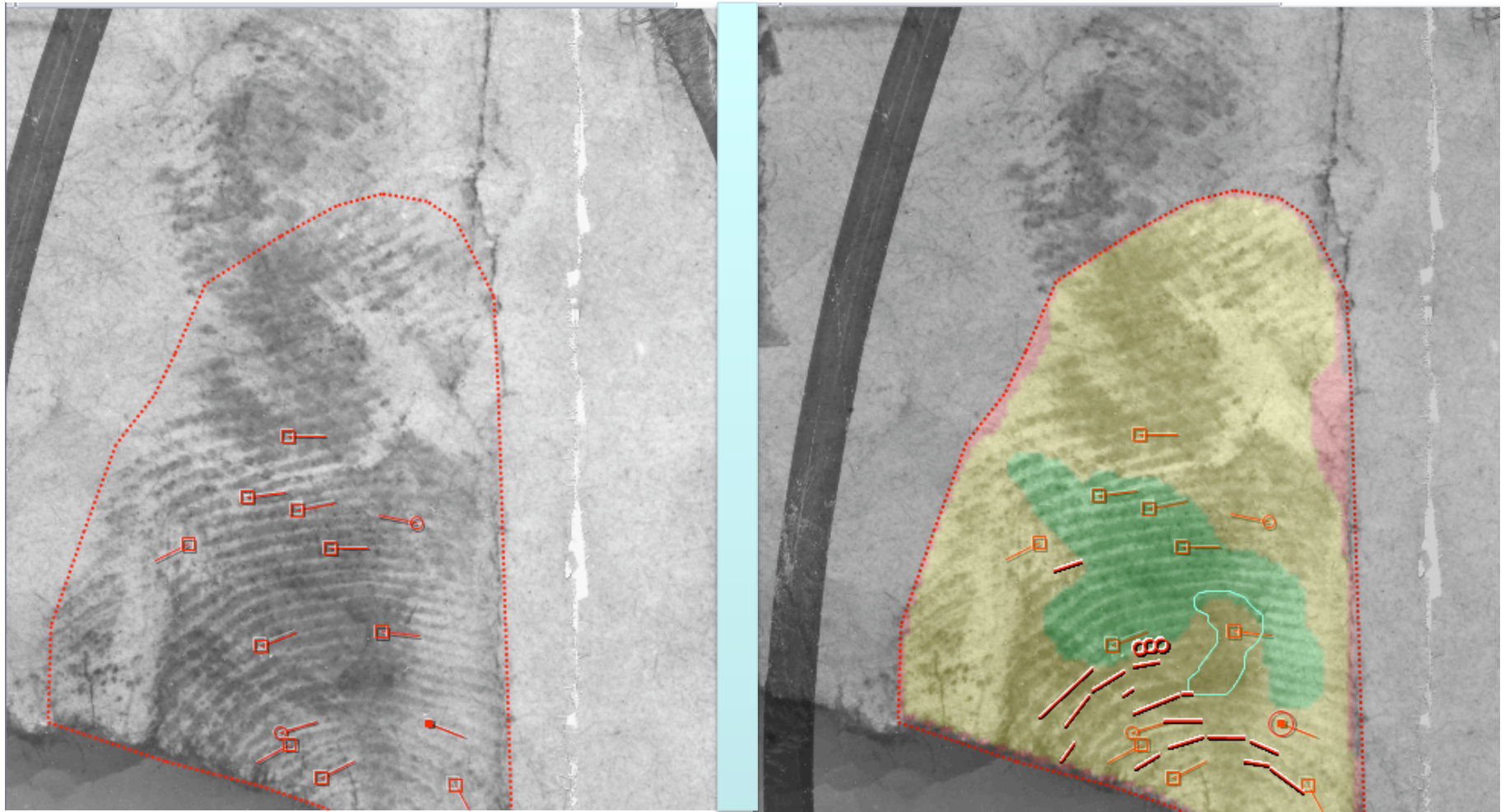
Fingerprint Feature Sets (continued...)

➤ Extended Feature Sets

- A “brave new standard” (updates ANSI/NIST)
- Enhanced handling of traditional features
- Larger “vocabulary” of features (e.g. “level 3”)
- Includes features already in proprietary sets to achieve `greater interoperability & performance
- ... also archivable: inter-examiner, legal evidence
- ... ***but will require testing before rollout***







Candidate List Workload Reduction

Rank	Subject	Similarity	Prob %
1	S709731	2903	92
2	S091303	2500	95
3	S213950	1754	40
4	S019893	1502	15
5	S400911	1245	5
...			
20	S004121	490	1

- ◆ Is the “mate” ranked near the top?
- ◆ Is “thresholding” effective?
- ◆ Can candidate lists be combined?

Evaluation of Latent Fingerprint Technologies (ELFT)

- Evaluation of Latent AFIS Performance:
 1. Image only searches (AFEM, “lights out”)
 2. Manually-assisted searches
 3. Measure accuracy, errors, speed (workload reduction capabilities)

Evaluation of Latent Fingerprint Technologies (ELFT)

- Two NIST-run tests: 24 AFEM technologies
- One self-run “challenge” : 6 AFEM + 5 EFS
- 3rd NIST-run test running... (5 AFEM + 5 EFS)

Future Directions

- Reverse latent searches (watchlist, ULF)
 - Latents vs. ID flats
 - Optimal thresholding strategies
- Latent quality metrics
 - quality directed workflow: AFEM or manual selection

For More Information...

Web → <http://fingerprint.nist.gov/latent>

Email → latent@nist.gov

What is the Problem?

- Highly specialized (niche) algorithms
 - Lack of independent performance evaluation
- Workload, Workload, Workload...
 - Extensive manual processing (pre & post search)
- Interoperability
 - Current systems use image-only, or ANSI/NIST defined features as a *framework* for interoperability
(variation limits interop; proprietary features perform better)

Standards

- Standardized enhanced features (CDEFFS)
 - building on ANSI/NIST-ITL 1-2007
 - feature-level interoperability
 - enhanced usage of traditional features
 - expanded feature set (e.g. “level 3” features)
- Image quality
 - development of quality metric for latents
 - enable quality directed workflow
- Testing and Evaluation
 - interfaces / protocols / metrics
- Standard Reference Data

2009 National Academies Report: “Strengthening Forensic Science in the United States: A Path Forward”

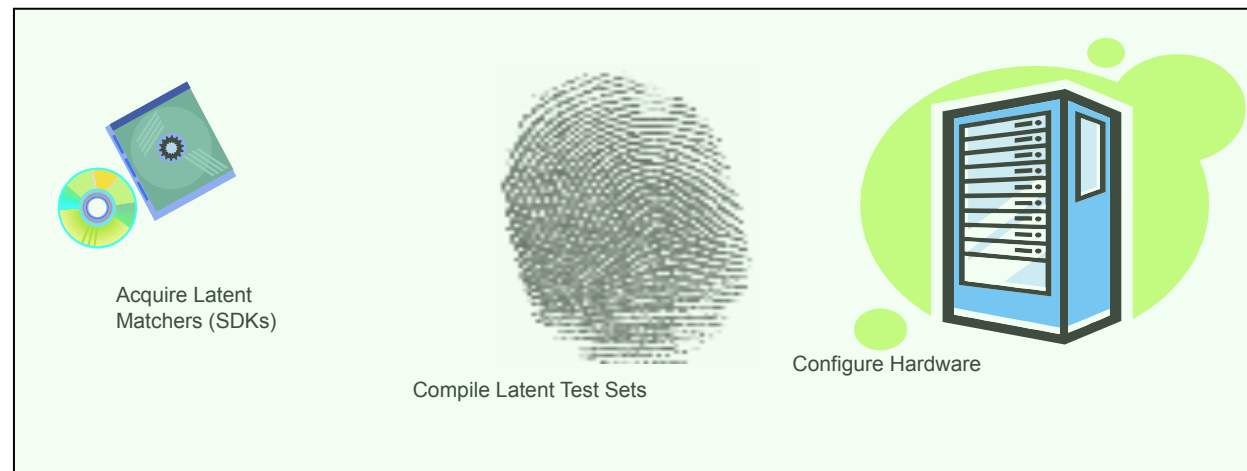
#3) “Research is needed to address issues of **accuracy**...”

#6) “...to develop tools for **advancing measurement, validation, reliability, information sharing**....”

#12) “...broad-based effort to achieve nationwide fingerprint data **interoperability** ... (a) **standards for representing and communicating image and minutiae data** among Automated Fingerprint Identification Systems; (b) baseline standards—to be used with **computer algorithms—to map, record, and recognize features in fingerprint images**, and a research agenda for the continued improvement, refinement, and **characterization of the accuracy of these algorithms (including quantification of error rates)**.”

ELFT Approach

Latent Testbed



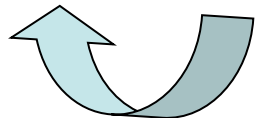
Evaluation Protocol

- 1:N matching tests
- “Open Universe”
- Multiple Resolutions
- Multiple Background Sizes
- Multiple Impression Types
- Multiple Data Sources
- Examiner Reviews

Evaluation Outputs

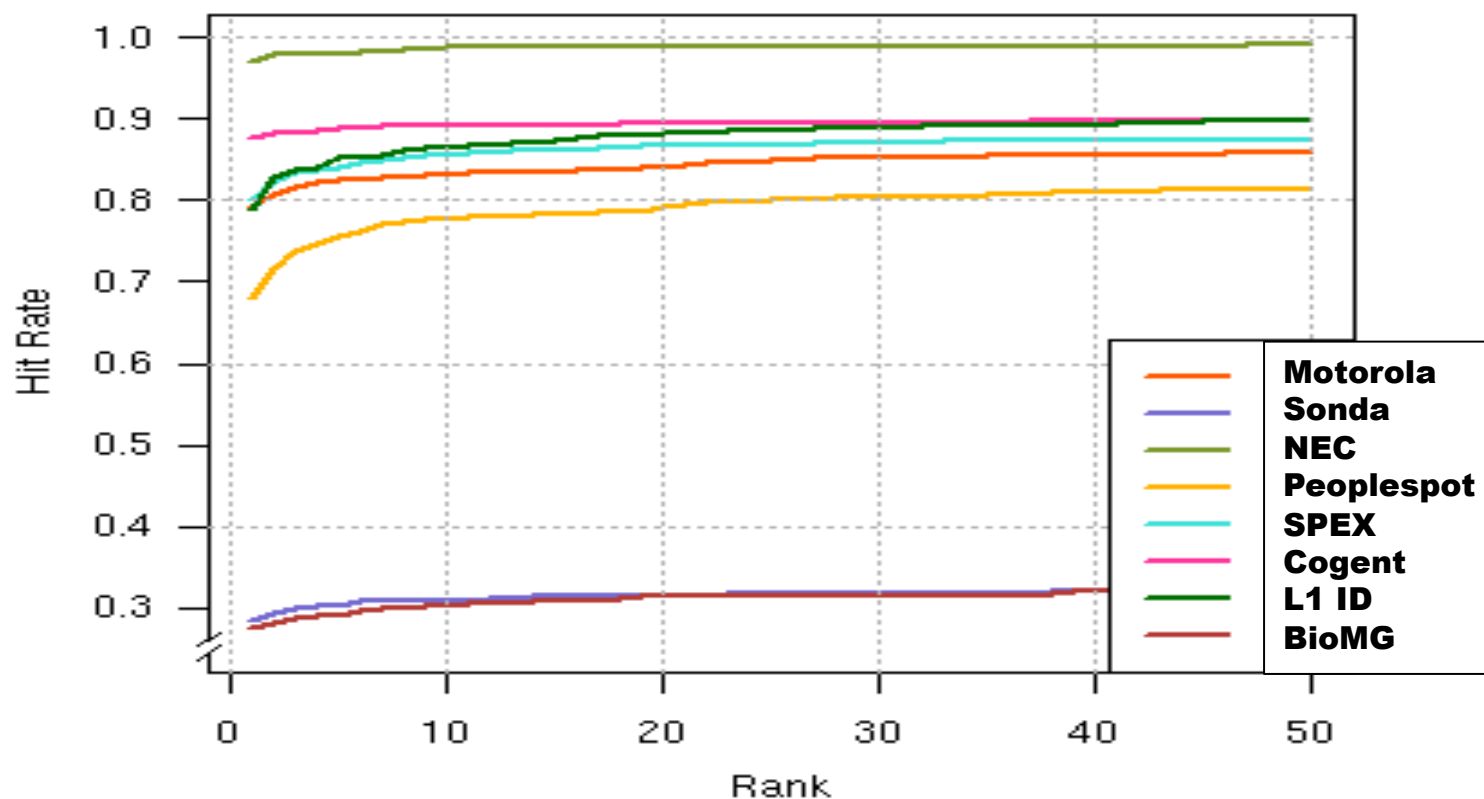
- Analytical Reports
- Reference Data/Miss analysis
- Updated Test Protocol
- Updated CONOPS & API
- Announcement of Phase N+1

Iterate process



Cumulative Match Characteristic (CMC)

1000 ppi latents • 100,000 fingerprint background



ELFT Relationship to NAS Report

➤ **Evaluation and testing of EFS promotes forensic fingerprint feature exchange by**

- Enabling standard recording of a latent examiner's casework
- Enabling standard presentation of a latent examiner's casework in court
- Between latent examiners (human-to-human)
- Between latent examiners and AFIS (human-to-computer)
- Between AFIS to AFIS (computer-to-computer)

➤ **Evaluation and testing of AFEM promotes improved computer algorithm mapping, recording, and recognition of of feature in fingerprint images**

Note: AFEM is complementary to the use of features as it may be used in place of or in addition to human extracted features

➤ **The primary objective of ELFT is to measure and characterize algorithmic accuracy and performance (including quantification of error rates).**

Latent Testing is *Different*

- **Scarcity of test data! (1000s vs. 1,000,000s)**
- Ground truth challenges
 - latents are collected not “captured”
 - latent to mate relationship must be established (typically using AFIS)
- 1:N evaluation protocols & metrics not mature
- Computational Complexity (8 SDKs/48 blades = 3 months)

ELFT Phase II Overview

- Tested 8 SDK's (one per participant), using
- Operational images from successful **feature** searches (IAFIS)
- Executed **image-only** searches using these images to measure overall AFEM accuracy
- Evaluated efficacy of candidate list reduction

Phase II Dataset

- Casework over 2 to 3 year period
 - Photographs (1000ppi) of developed prints
 - Acquired from paper sources
- 835 images from 588 unique subjects
 - additional 500ppi (sub-sampled) set
 - additional set of Region-of Interest (ROI) overlays
- Database of 100,000 fingerprints
 - mixed operational sources (4 sources; civil & criminal)
 - mixed types (inked and live-scanned)

Phase II Testing Protocol

Latent Image Resolution (ppi)	Database Size (fingerprints)	ROI overlay?
1000	100,000	No
1000	50,000	No
500	50,000	No
1000	50,000	Yes

- ✓ Overall accuracy; generate data for error analysis
- ✓ Effect of database size (scalability)
- ✓ Effect of latent image resolution
- ✓ Effect of Region-of-Interest (ROI) markup

Metrics

- **Rank-based**

- Proportion of searches resulting in the true-match (“mate”) appearing on the candidate list
- Position (“ranks”) at which mate appears

- **Threshold-based**

- Proportion of searches resulting in false-matches (“non-mates”) appearing on the candidate list above a “score threshold”
- Accuracy (identification rate) after thresholding
(*can we screen out false-matches without losing too many idents?*)

Phase II Analysis Results

- **Overall Accuracy**
- **Effect of Database Size**
- **Effect of Resolution**
- **Effect of ROI**
- **Effect of Minutiae Count**
- **Candidate List Fusion (multi-image & multi-algorithm)**
- **Effect of Finger Position**
- **Effect of Pattern Class**
- **Execution Times**

Detection Rates (**Rank 1**)

1000 ppi latents • 100,000 fingerprint background

SDK	Technology Provider	Detection Rate at Rank 1
M1	NEC	97.2
P1	Cogent	87.8
O1	SPEX	80.0
K1	Motorola	79.3
Q1	L1 Identity Solutions	78.8
N1	Peoplespot	67.9
L1	Sonda	28.5
R1	BioMG	27.5

$\mu = 82\%$

Detection Rates (**Rank 10**)

1000 ppi latents • 100,000 fingerprint background

SDK	Technology Provider	Detection Rate at Rank 10
M1	NEC	98.8
P1	Cogent	89.2
Q1	L1 Identity Solutions	86.5
O1	SPEX	85.6
K1	Motorola	83.2
N1	Peoplespot	77.8
L1	Sonda	30.9
R1	BioMG	30.2

◀ +1.6

◀ +1.4

◀ +7.7

◀ +5.6

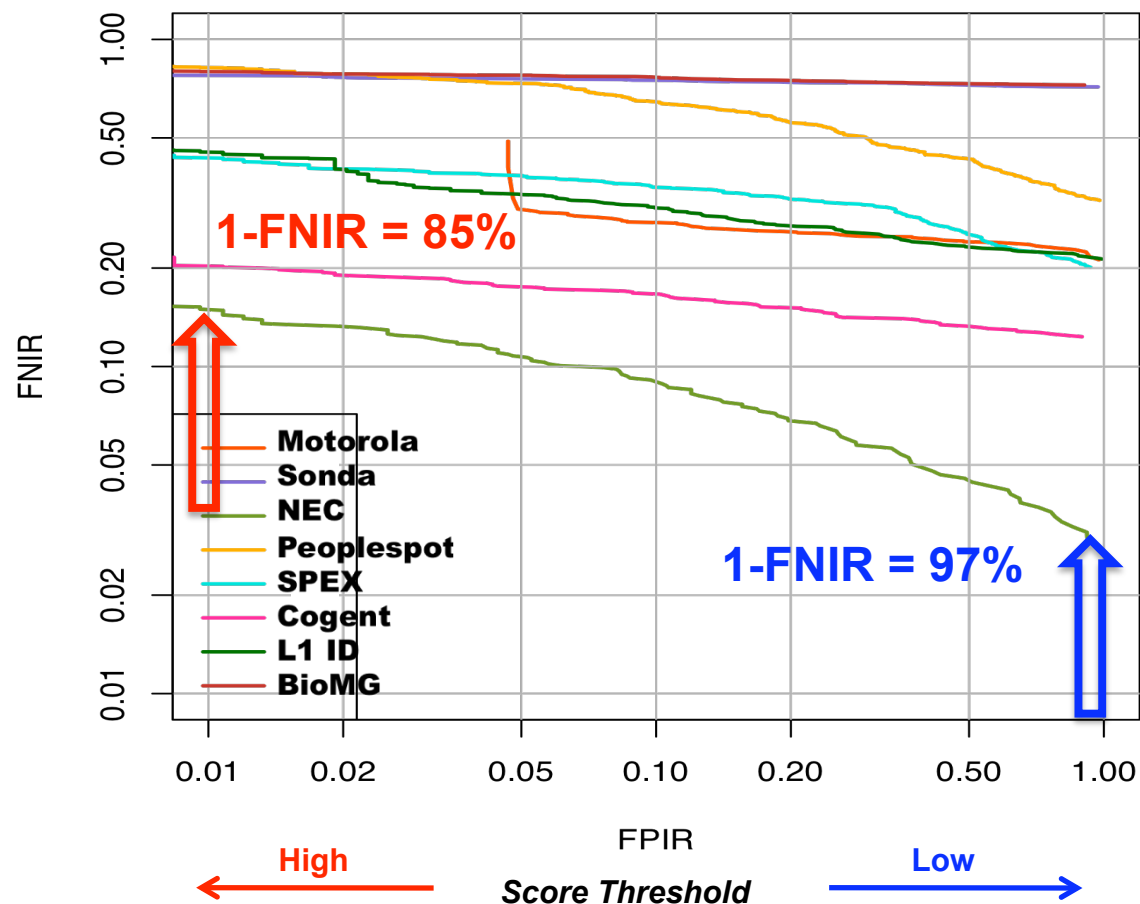
◀ +3.9

◀ +9.9

 $\mu = 87\%$

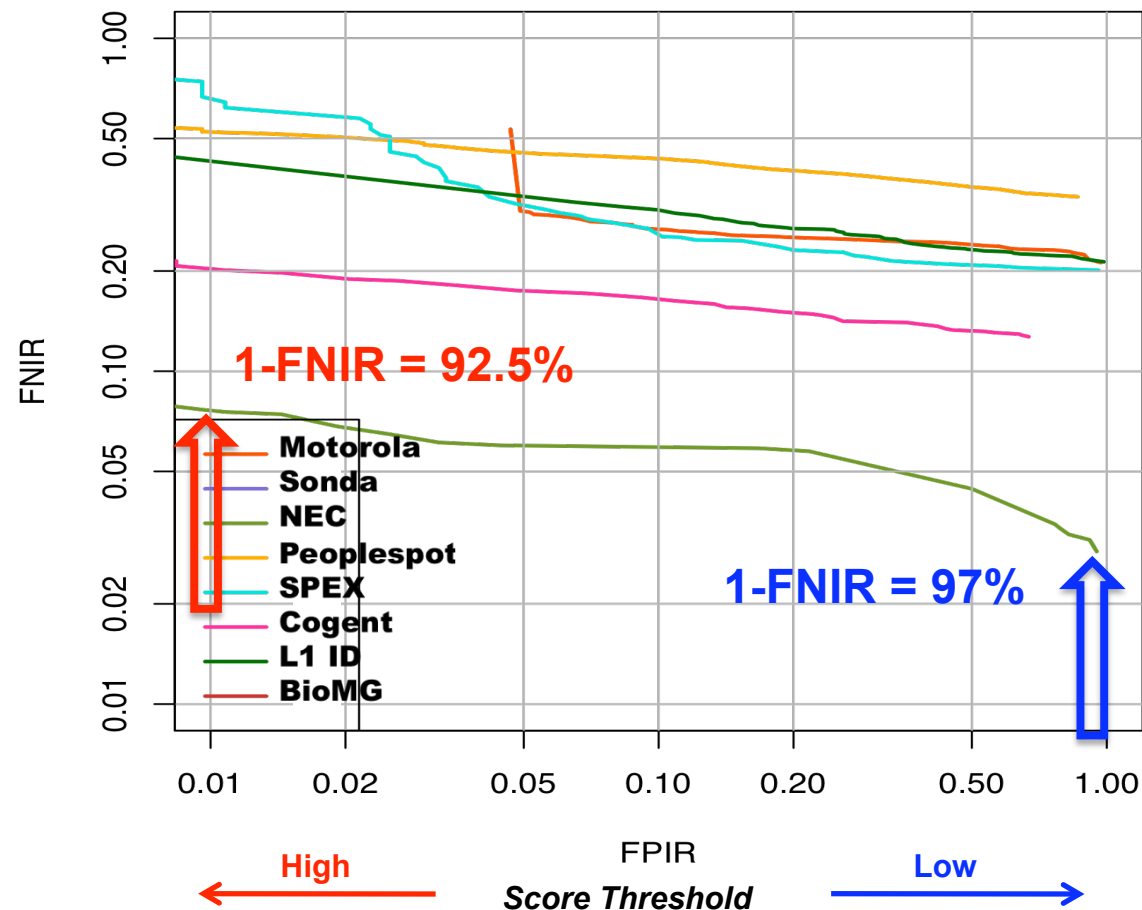
DET – *match score*

1000 ppi latents • 100,000 fingerprint background



DET – *probability score*

1000 ppi latents • 100,000 fingerprint background



Phase II Observations - 1

- *Thresholding* based on an SDK provided *probability* score was shown to be more effective at reducing false-matches than the provided proprietary scores for two SDKs. This has important implications for candidate list reduction, interoperability, and fusion.
- A strong correlation exists between minutiae count and identification rate. Searches of latents with higher minutiae counts produced more accurate results.
- Candidate list fusion, using multi-fingers or multi-algorithms is a powerful mechanism for improving accuracy.

Phase II Observations - 2

- The effect of increasing database size from 50,000 to 100,000 resulted in a one percentage-point average decrease in accuracy at rank 1
- The effect of resolution (1000 ppi vs. 500 ppi) was mixed and not statistically significant.
- The effect of region-of-interest itself was mixed, however, images with >50% area of ROI benefited the most.

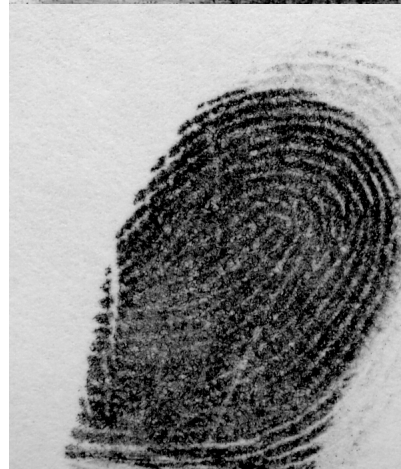
Phase II Conclusions

1. Some matchers tested possess accuracies such that a limited class of latent fingerprints from operational casework can benefit from AFEM, thereby reducing some of the human workload during the AFIS latent fingerprint processes.
2. Specific measures (e.g. latent quality measures) do not currently exist for determining which latents are suitable for AFEM.
3. More testing is required to define AFEM limitations.

Phase II Caveats

- Participants were encouraged to submit research algorithms which may not be in operational use or commercially available
- Latents and exemplars were identified by an operational AFIS
- AFEM accuracy is highly dependent on source, selection, and preparation of data. Study results may not be applicable to other datasets and operational databases.
- Minimal constraints on processing time (not necessarily reflective of operational requirements)

What are Latents?



Stakeholders

➤ Law Enforcement: Federal, Local, and International

➤ Counter-terrorism / Counter-insurgency

E.g.)

- ◆ US-VISIT's IDENT latent watchlist contains ~40,000 prints
 - unidentified latents from FBI, DoD, intelligence agencies, etc.
- ◆ Captured prints from 100,000 subject's searched against watchlist
 - prints captured at ports of entry, embassies, and consulates
- ◆ 18% searches result in "candidates" (**manually verified**)
 - high false-alarm rate (vast majority of are non-idents)
 - low hit rate (367 subjects identified in 2008)
 - 41 staff examiners make about 2,000 negative idents weekly